Statistique Mathématique

Salim Lardjane

Université Bretagne Sud

Partie III

Estimateurs du Maximum de Vraisemblance, des Moments et Intervalles de Confiance

Estimateurs du Maximum de Vraisemblance (EMV)

Dans ce cas $\delta(x) = \widehat{\theta}$, une *valeur possible* de θ pour laquelle ce que l'on a observé était le *plus vraisemblable* possible.

En d'autres termes, $\widehat{\theta}$ maximise la *fonction de vraisemblance* :

$$L(\theta) = f_\theta(x) \quad (x \text{ fixé}).$$

C'est donc une fonction de x uniquement, comme il se doit pour que l'estimation ait un sens.

On note généralement de façon identique *l'estimation* $\widehat{\theta}$, fonction de x, et *l'estimateur* $\widehat{\theta}$, fonction de X.

Estimateurs du Maximum de Vraisemblance (EMV)

Exemple 1. Supposons que $X \sim \mathcal{B}(n, \theta)$. Alors,

$$\mathbb{P}_\theta(X = x) = f_\theta(x) = C_n^x \theta^x (1-\theta)^{n-x}$$

pour $x = 0, 1, \ldots, n$.

Maximiser $L(\theta) = f_\theta(x)$ revient à maximiser

$$\ell(\theta) = \log L(\theta).$$

Ecrivons la condition au premier ordre :

$$\begin{aligned} \frac{d\ell(\theta)}{d\theta} &= \frac{d}{d\theta}(x \log \theta + (n-x)\log(1-\theta)) \\ &= \frac{x}{\theta} - \frac{n-x}{1-\theta} = 0 \end{aligned}$$

On en déduit $x - \theta x - \theta n + \theta x = 0$, d'où

$$\widehat{\theta} = \frac{x}{n}.$$

Estimateurs du Maximum de Vraisemblance (EMV)

Ainsi, l'estimation de θ par maximum de vraisemblance dans le cas binomial correspond à la fréquence relative de succès.

On note :

$$\widehat{\theta}_n = \frac{X}{n}$$

l'*estimateur* associé.

Puisque $X \sim \mathcal{B}(n, \theta)$, on a de façon immédiate :

$$\mathbb{E}(\widehat{\theta}_n) = \frac{n\theta}{n} = \theta.$$

Par conséquent, l'EMV $\widehat{\theta}_n$ est *sans biais*.

De plus, en vertu de la Loi Faible des Grands Nombres, on a :

$$\widehat{\theta}_n \to_{\mathbb{P}} \theta \quad (n \to \infty)$$

Ainsi, l'EMV $\widehat{\theta}_n$ est *consistant*.

Estimateurs du Maximum de Vraisemblance (EMV)

Exemple 2. Considérons le cas où $X_1, \dots, X_n$ sont i.i.d. de loi $\mathcal{N}(\mu, \sigma^2)$.

Ici, $\theta = (\mu, \sigma)$ est un paramètre *vectoriel*.

En notant $x = (x_1, \dots, x_n)$ une *réalisation* de l'échantillon $X_1, \dots, X_n$, on peut écrire la vraisemblance :

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{n} f_\theta(x_i) \quad \text{(indépendance)} \\ &= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left(-\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}\right) \end{aligned}$$

d'où

$$\begin{aligned} \ell(\theta) &= \log L(\theta) \\ &= -\frac{n}{2}\log(2\pi) - n\log(\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 \end{aligned}$$

Estimateurs du Maximum de Vraisemblance (EMV)

Au maximum, on a les conditions au premier ordre :

$$\frac{\partial \ell(\theta)}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu) = 0$$

d'où $\mu = \bar{x}$, et

$$\begin{aligned} \frac{\partial \ell(\theta)}{\partial \sigma} &= -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^{n} (x_i - \mu)^2 \\ &= \frac{n}{\sigma^3}(-\sigma^2 + \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2) \end{aligned}$$

avec $\mu = \bar{x}$. Ainsi, au maximum, $\mu = \bar{x}$ et

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 = s^2.$$

Estimateurs du Maximum de Vraisemblance (EMV)

Le calcul précédent est valable lorsque μ et σ^2 sont *inconnus*. Dans ce cas, l'EMV est donné par

$$\widehat{\theta}_n = (\bar{X}, S^2).$$

Dans le cas où σ^2 est connu et μ inconnu, un calcul analogue donne

$$\widehat{\theta}_n = \bar{X}$$

puisque dans ce cas $\theta = \mu$.

Dans le cas où μ est connu et σ^2 inconnu, on obtient :

$$\widehat{\theta}_n = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2.$$

Estimateurs du Maximum de Vraisemblance (EMV)

La moyenne arithmétique $\bar{X}$ est un estimateur sans biais et, en vertu de la Loi Faible des Grand Nombres, consistant de μ.

La variance empirique S^2 est un estimateur biaisé mais consistant de la variance théorique σ^2 (vu précédemment).

Remarque. En toute rigueur, il faut calculer les *dérivées secondes* pour s'assurer que l'extremum obtenu est bien un *maximum*.

Estimateurs du Maximum de Vraisemblance (EMV)

Une propriété très utile de l'EMV est la suivante. Soit h une fonction de θ à valeurs réelles, alors :

$$\widehat{h(\theta)} = h(\widehat{\theta}).$$

De plus, si h est *continue*, *la consistance est préservée*, c'est-à-dire que

$$h(\widehat{\theta}_n) \to_{\mathbb{P}} h(\theta)$$

dès lors que $\widehat{\theta}_n \to_{\mathbb{P}} \theta$.

C'est une conséquence des théorèmes de Slutsky (p. 34).

Méthode des Moments

Une méthode parfois plus rapide que l'EMV pour déterminer une estimation raisonnable du paramètre θ est la *méthode des moments*.

Elle consiste à égaler le k-ème *moment empirique* :

$$\frac{1}{n}\sum_{i=1}^{n} X_i^k$$

au k-ème moment théorique $\mathbb{E}(X_i^k)$, qui dépend du paramètre inconnu θ.

On considère autant d'équations de moments qu'il est nécessaire pour déterminer une estimation θ^*.

Méthode des Moments

Alternativement, ont peut égaler le k-ème moment empirique *centré* :

$$\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^k$$

au k-ème moment centré théorique $\mathbb{E}[(X_i - \mu)^k]$.

Exemple. Soient $X_1, \ldots, X_n$ i.i.d., de loi gamma $\gamma(\theta_1, \theta_2)$, avec $\theta_1, \theta_2 > 0$.

Alors, on a vu précédemment que :

$$\mathbb{E}(X_i) = \theta_1\theta_2, \quad V(X_i) = \theta_1\theta_2^2$$

Méthode des Moments

On pose alors le système d'équations :

$$\begin{cases} \bar{X} & = & \theta_1\theta_1 \\ S^2 & = & \theta_1\theta_2^2 \end{cases}$$

et on résout pour obtenir des estimateurs θ_i^* de θ_i $(i = 1, 2)$.

Plus précisément, on obtient :

$$\theta_2^* = \frac{S^2}{\bar{X}}$$

$$\theta_1^* = \frac{\bar{X}^2}{S^2}.$$

Intervalles de Confiance

Supposons qu'il y ait deux candidats A et B à une élection.

La probabilité qu'un électeur préfère A à B est notée p.

p est une *constante inconnue*

On sélectionne n électeurs indépendamment et on leur demande pour qui ils vont voter.

Intervalles de Confiance

Le nombre Y_n d'électeurs qui votent A suit une loi binomiale $\mathcal{B}(n, p)$ qui, pour n grand, peut être approximée par une loi normale d'espérance $\mu = np$ et de variance $\sigma^2 = np(1-p)$.

La fréquence relative, ou *proportion*, d'électeurs qui votent A dans l'échantillon est $\frac{Y_n}{n}$.

Intervalles de Confiance

On veut déterminer la valeur minimale de n telle qu'on puisse prédire la part du vote en faveur de A à 1% près, *au niveau de confiance* 95%, c'est-à-dire telle que :

$$\mathbb{P}\left(\left|\frac{Y_n}{n} - p\right| \leq 0.01\right) \geq 0.95$$

ou encore :

$$\mathbb{P}\left(\left[\frac{Y_n}{n} - 0.01, \frac{Y_n}{n} + 0.01\right] \ni p\right) \geq 0.95.$$

Intervalles de Confiance

Autrement dit, on souhaite que l'intervalle *aléatoire* :

$$I_n = \left[\frac{Y_n}{n} - 0.01, \frac{Y_n}{n} + 0.01\right]$$

contienne la valeur inconnue de p avec une probabilité supérieure ou égale à 0.95.

On dit que I_n est *un intervalle de confiance à 95%* pour p.

Intervalles de Confiance

On peut écrire :

$$\begin{aligned}
\mathbb{P}\left(\left|\frac{Y_n}{n} - p\right| \leq 0.01\right) &= \mathbb{P}(|Y_n - np| \leq 0.01\, n) \\
&= \mathbb{P}\left(\left|\frac{Y_n - np}{\sqrt{np(1-p)}}\right| \leq \frac{0.01\,\sqrt{n}}{\sqrt{p(1-p)}}\right) \\
&\approx \Phi\left(\frac{0.01\,\sqrt{n}}{\sqrt{p(1-p)}}\right) - \phi\left(\frac{-0.01\,\sqrt{n}}{\sqrt{p(1-p)}}\right) \\
&= 2\Phi\left(\frac{0.01\,\sqrt{n}}{\sqrt{p(1-p)}}\right) - 1 \geq 0.95
\end{aligned}$$

où Φ est la fonction de répartition de la loi normale standard $\mathcal{N}(0,1)$.

Intervalles de Confiance

Comme 1.95/2=0.975 et $\Phi(1.96) = 0.975$, on en déduit :

$$\frac{0.01\,\sqrt{n}}{\sqrt{p(1-p)}} \geq 1.96$$

d'où

$$n \geq 196^2\, p(1-p)$$

Or $p(1-p)$ est maximale lorsque $p = 1/2$ et vaut dans ce cas 1/4.

On doit donc avoir, pour couvrir tous les cas possibles :

$$n \geq 196^2/4 = 9604$$

Ainsi, si on veut connaître p à 1% près au niveau de confiance 95%, il faut une taille d'échantillon supérieure ou égale à 9604 individus.

Si on veut connaître p à 1% près au niveau de confiance 99%, on procède de la même façon et on obtient qu'on doit avoir :

$$n \geq 1\,690\,000$$

Intervalles de Confiance

Si on veut connaître p à 3% près au niveau de confiance 95%, on montre de même qu'on doit avoir :

$$n \geq 1067$$

Il est remarquable que la précision de la prédiction ne dépend que du nombre d'électeurs interrogés et pas du nombre total d'électeurs.

Intervalles de Confiance

A présent, soit $X_1, \ldots, X_n$ un échantillon i.i.d. gaussien associé à la loi $\mathcal{N}(\mu, \sigma^2$.

On va déterminer *un intervalle de confiance pour* μ.

A cet effet, on va distinguer deux cas.

Cas 1 : La variance σ^2 est connue.

Dans ce cas, $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$ et

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1).$$

Par conséquent, pour $b > 0$, on a :

$$\begin{aligned} \mathbb{P}\left(-b \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq b\right) &= \Phi(b) - \Phi(-b) \\ &= 2\Phi(b) - 1. \end{aligned}$$

Intervalles de Confiance

Autrement dit :

$$\mathbb{P}\left(\left[\bar{X} - \frac{b\sigma}{\sqrt{n}}, \bar{X} + \frac{b\sigma}{\sqrt{n}}\right] \ni \mu\right) = 2\Phi(b) - 1$$

En fixant le niveau de confiance, on détermine b, qui à son tour détermine l'intervalle de confiance.

Par exemple, pour un niveau de confiance de 95%, on a $b = 1.96$.

Cas 2 : La variance σ^2 est inconnue.

Dans ce cas, on a vu précédemment que :

$$\frac{\bar{X} - \mu}{S/\sqrt{n-1}} \sim T_{n-1}.$$

Par conséquent, pour $b > 0$, on a :

$$\mathbb{P}\left(-b \leq \frac{\bar{X} - \mu}{S/\sqrt{n-1}} \leq b\right) = 2F_{T_{n-1}}(b) - 1.$$

On détermine b à partir du niveau de confiance et des tables de la loi de Student à $n-1$ degrés de liberté, puis on en déduit l'intervalle de confiance.

IC pour une différence de moyennes

Soit $X_1, \ldots, X_n$ un échantillon i.i.d. associé à une v.a.r. X de loi $\mathcal{N}(\mu_1, \sigma_1^2)$ et soit $Y_1, \ldots, Y_m$ un échantillon i.i.d. associé à une v.a.r. Y de loi $\mathcal{N}(\mu_2, \sigma_2^2)$.

On suppose d'abord que $\sigma_1^2 = \sigma_2^2 = \sigma^2$, quantité *inconnue*.

On suppose de plus que $(X_1, \ldots, X_n)$ et $(Y_1, \ldots, Y_m)$ sont indépendants.

On souhaite construire un intervalle de confiance pour $\mu_1 - \mu_2$.

Dans la pratique, l'intervalle obtenu est souvent utilisé de la façon suivante : *si la borne supérieure de l'intervalle est inférieure à 0, on a de bonnes raisons de croire que* $\mu_1 < \mu_2$.

IC pour une différence de moyennes

Comme $V(\bar{X}-\bar{Y}) = V(\bar{X}) + V(\bar{Y}) = \sigma^2/n + \sigma^2/m$, on a :

$$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}} \sim \mathcal{N}(0,1).$$

De plus, $nS_1^2/\sigma^2 \sim \chi^2(n-1)$, $mS_2^2/\sigma^2 \sim \chi^2(m-1)$ et, en vertu de la définition de la loi du Khi-deux :

$$\frac{nS_1^2}{\sigma^2} + \frac{mS_2^2}{\sigma^2} \sim \chi^2(n+m-2).$$

IC pour une différence de moyennes

Notons :

$$R = \sqrt{\left(\frac{nS_1^2 + mS_2^2}{n+m-2}\right)\left(\frac{1}{n} + \frac{1}{m}\right)}$$

Alors :

$$T = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{R} \sim St_{n+m-2}.$$

L'hypothèse que les deux populations ont même variance σ^2 est cruciale car on peut de ce fait l'éliminer.

IC pour une différence de moyennes

En écrivant :

$$\mathbb{P}(-b \leq T \leq b) = 0.95,$$

on obtient un intervalle de confiance de niveau 95% pour $\mu_1 - \mu_2$:

$$-b \leq \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{R} \leq b$$

autrement dit :

$$\bar{X} - \bar{Y} - bR \leq \mu_1 - \mu_2 \leq \bar{X} - \bar{Y} + bR,$$

où b est le fractile d'ordre 0.975 de la loi St_{n+m-2}.

IC pour une différence de moyennes

Si les variances σ_1^2 et σ_2^2 sont *connues* et éventuellement *inégales*, alors :

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \sim \mathcal{N}(0, 1).$$

Notons :

$$R_0 = \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}.$$

Alors, un intervalle de confiance de niveau 95% pour $\mu_1 - \mu_2$ est donné par :

$$\bar{X} - \bar{Y} - bR_0 \leq \mu_1 - \mu_2 \leq \bar{X} - \bar{Y} + bR_0,$$

où b désigne le fractile d'ordre 0.975 de la loi $\mathcal{N}(0, 1)$

IC pour une différence de moyennes

Exemple complémentaire. Soient $Y_1 \sim \mathcal{B}(n_1, p_1)$ et $Y_2 \sim \mathcal{B}(n_2, p_2)$. Alors, on peut écrire :

$$Y_1 = X_1 + \cdots + X_{n_1} \quad \text{et} \quad Y_2 = Z_1 + \cdots + Z_{n_2}$$

où les $X_i \sim \mathcal{B}(p_1)$ $(i = 1, \ldots, n_1)$ sont des variables de Bernoulli indépendantes et les $Z_i \sim \mathcal{B}(p_2)$ $(i = 1, \ldots, n_2)$ également.

Supposons de plus que $(X_1, \ldots, X_{n_1})$ et $(Z_1, \ldots, Z_{n_2})$ sont indépendants.

Nous avons :

$$\mathbb{E}(Y_1/n_1) = p_1, \quad V(Y_1/n_1) = p_1(1 - p_1)/n_1$$

et des formules analogues pour Y_2.

IC pour une différence de moyennes

Par conséquent, pour n grand :

$$\left(\frac{Y_1}{n_1} - \frac{Y_1}{n_2}\right) - (p_1 - p_2)$$

divisé par :

$$\sqrt{\frac{p_1(1-p_1)}{1} + \frac{p_2(1-p_2)}{n_2}}$$

suit approximativement la loi $\mathcal{N}(0,1)$.

Mais cette expression *ne peut être utilisée pour construire des intervalles de confiance pour* $p_1 - p_2$ car le dénominateur fait intervenir les quantités *inconnues* p_1 et p_2.

Toutefois, Y_1/n_1 converge en probabilité vers p_1 et Y_2/n_2 converge en probabilité vers p_2.

On remplace donc, en pratique, p_1 par Y_1/n_1 et p_2 par Y_2/n_2 dans le déominateur.

IC pour la variance

Soit $X_1, \ldots, X_n$ un échantillon i.i.d. associé à une v.a.r. X de loi $\mathcal{N}(\mu, \sigma^2)$ où μ et σ^2 sont inconnues.

Alors :

$$\frac{nS^2}{\sigma^2} \sim \chi^2(n-1)$$

Notons $h_{n-1}(x)$ la densité de la loi $\chi^2(n-1)$ et soient a et b tels que :

$$\int_a^b h_{n-1}(x)dx = 1 - \alpha.$$

IC pour la variance

Alors :

$$\mathbb{P}(a \leq \frac{nS^2}{\sigma^2} \leq b) = 1 - \alpha,$$

c'est-à-dire que :

$$\frac{nS^2}{b} \leq \sigma^2 \leq \frac{nS^2}{a}$$

est un intervalle de confiance de niveau $1 - \alpha$ pour σ^2.

IC pour la variance

En pratique, a et b sont déterminés de façon à vérifier :

$$\int_b^{\infty} h_{n-1}(x)dx = \int_{-\infty}^{a} h_{n-1}(x)dx.$$

Par exemple, en notant $H_{n-1}(x)$ la fonction de répartition de la loi $\chi^2(n-1)$ et pour $\alpha = 0.05$, on pose :

$$\begin{aligned} H_{n-1}(a) &= 0.025 \\ H_{n-1}(b) &= 1 - 0.025 = 0.975 \end{aligned}$$

ce qui permet de déterminer a et b.

IC pour la variance

Remarque. Dans le développement précédent, μ est inconnue. Si μ est connue, on peut exploiter ce fait pour obtenir un meilleur intervalle de confiance.

En effet, nous avons alors :

$$\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2 \sim \chi^2(n).$$

Notons :

$$W = \sum_{i=1}^{n} (X_i - \mu)^2$$

et soient a et b tels que $\int_a^b h_n(x)dx = 1 - \alpha$, où $h_n(x)$ désigne la densité de la loi $\chi^2(n)$.

IC pour la variance

Alors :

$$\mathbb{P}(a \leq \frac{W}{\sigma^2} \leq b) = 1 - \alpha,$$

d'où l'intervalle de confiance de niveau $1 - \alpha$ pour σ^2 :

$$\frac{W}{b} \leq \sigma^2 \leq \frac{W}{a}.$$

IC pour un rapport de variances

Soit $X_1, \ldots, X_n$ un échantillon i.i.d. associé à une v.a.r. X de loi $\mathcal{N}(\mu_1, \sigma_1^2)$ et soit $Y_1, \ldots, Y_m$ un échantillon i.i.d. associé à une v.a.r. Y de loi $\mathcal{N}(\mu_2, \sigma_2^2)$.

On suppose de plus que $(X_1, \ldots, X_n)$ et $(Y_1, \ldots, Y_m)$ sont indépendants.

Alors :

$$\frac{n_i S_i^2}{\sigma_i^2} \sim \chi^2(n_i - 1) \quad (i = 1, 2).$$

Par conséquent :

$$\frac{(n_2 S_2^2/\sigma_2^2)/(n_2 - 1)}{(n_1 S_1^2/\sigma_1^2)/(n_1 - 1)} \sim F_{n_2-1, n_1-1}.$$

IC pour un rapport de variances

Mais :

$$V_i^2 = n_i S_i^2/(n_i - 1) \quad (i = 1, 2)$$

sont les estimateurs sans biais usuels de la variance et on a donc, d'après ce qui précède :

$$\frac{V_2^2\,\sigma_1^2}{V_1^2\,\sigma_2^2} \sim F_{n_2-1,n_1-1},$$

ce qui permet d'obtenir un intervalle de confiance pour σ_1^2/σ_2^2 de la façon habituelle.